CECA54A/CECS54A/CSC53 — DATA MINING

Time : Three hours   Maximum : 75 marks

SECTION A — (10 × 2 = 20 marks)

Answer ALL questions.

1.   What is data visualization?

2.   Define data transformation.

3.   What is a data cube?

4.   What is data integration?

5.   Define Multidimensional space.

6.   What is the colossal pattern?

7.   What is the use of partitioning?

8.   State the importance of grids.

9.   Write the purpose of WEKA tool.

10. Comment on Apache mahout.

SECTION B — (5 × 5 = 25 marks)

Answer ALL questions.

11. (a) Expound the various issues in data mining.

Or

(b) Write short notes on data transformation.

12. (a) Discuss the different types of datacube technology.

Or

(b) Comment on data Warehouse design and its usage.

13. (a) Give an account on pattern mining in multilevel.

Or

(b) Expound rule based classification.

14. (a) Elaborate clustering with high dimensional data.

Or

(b) Give a brief note on outlier detection method.

15. (a) Enumerate the salient features of Rapidminer.

Or

(b) Discuss the various trends in data mining.

SECTION C — (3 × 10 = 30 marks)

Answer any THREE questions.

16. Describe the following in detail :

(a) Data cleaning

(b) Data integration

(c) Data reduction

(d) Data discretization.

17. Bring out the importance of multidimensional data analysis in cube space.

18. Expound the techniques to improve classification accuracy.

19. Differentiate web mining with spatial mining.

20. What are the other methodologies of data mining? Explain.